Structural Analysis Quick Start
An NCBI Mini-Course

A protein domain is considered to be a distinct functional and/or structural unit. A domain in a structural context refers to a segment of a polypeptide chain that can fold into an independent three dimensional structure. It may interact with other domains of the protein or may simply be joined to other domains by a polypeptide chain. A domain in a sequence context refers to a long sequence pattern that is shared by other proteins having a common evolutionary origin. A domain may include all of the protein sequence or a part of it. A conserved domain is a recurring unit in molecular evolution whose extents can be determined by sequence and structure analysis.

The Conserved Domain Database (CDD) contains domains derived from the Smart, Pfam and Clusters of Orthologous Groups (COGs) databases. Conserved domains can be represented as multiple sequence alignments. Source alignments are processed by NCBI as follows:

- Sequences in the alignment for which a link can not be provided to a protein in Entrez are removed.
- If possible, a closely related sequence with a known structure is substituted.
- A representative sequence, preferably with a structure link, is chosen from among those in the alignment.
- A consensus sequence is made.
- A position-specific scoring matrix (PSSM) is constructed.

The Conserved Domain search (CD-search) compares a protein sequence to the PSSMs in the CDD database to identify conserved domains within it and to identify a 3-D modeling template. Since the PSSMs are the "subject", instead of the query as in PSI-Blast, the CD-search is a form of Reverse Position-Specific Blast (RPS-Blast).

The Conserved Domain Architecture Retrieval Tool (CDART) can be used to identify proteins containing the domain(s) present in the query sequence. Conserved domain(s) present in all sequences within Entrez proteins are identified using CD-search during routine NCBI processing. These pre-computed results are accessed through CDART.

The Vector Alignment Search Tool (VAST) is a computer algorithm developed at NCBI to detect similar protein 3-dimensional structures. The "structure neighbors" for every structure in NCBI' Molecular Modeling DataBase (MMDB)

are pre-computed. These neighbors can be used to identify distant homologs that cannot be recognized by sequence comparison alone. A VAST-search can be used for determining the structure neighbors for recently solved structures not yet in MMDB.

Cn3D is a helper application for web browsers to view 3-dimensional structures from NCBI's Entrez retrieval service. Cn3D runs on Windows, Macintosh, and Unix. Cn3D simultaneously displays structure, sequence, and alignment, and now has powerful annotation and alignment editing features.

In this course, we will learn to

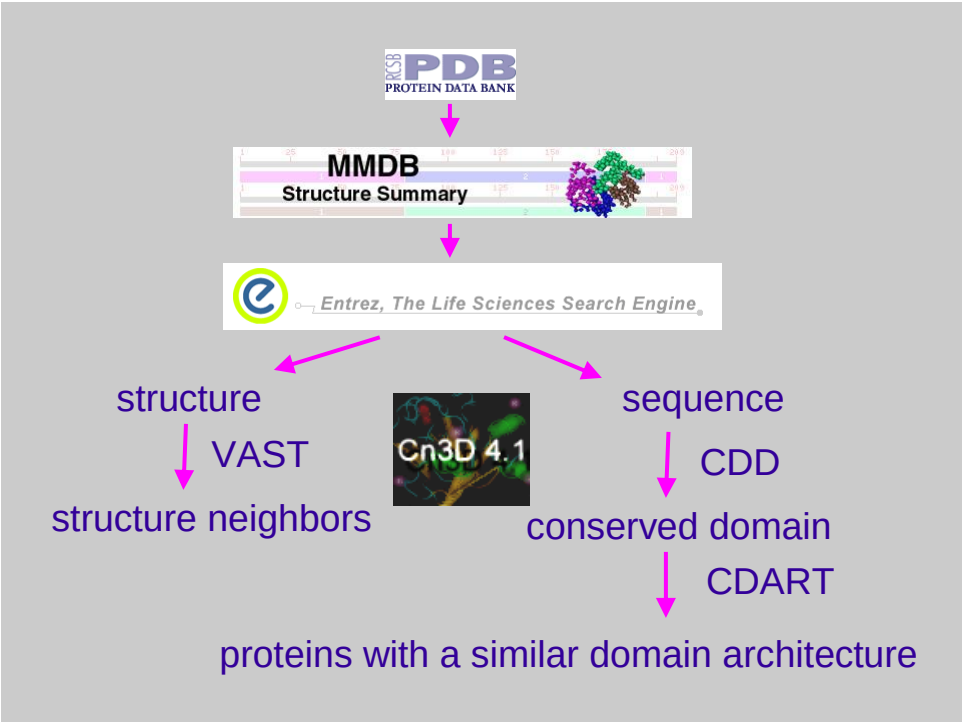
- Identify a conserved domain present in the query protein using **CDD**
- Search for other proteins containing similar domain(s) using **CDART**
- Explore a 3D modeling template for the query sequence using **CDD**
- Find similar structures using **VAST**
- Visualize and annotate the 3D protein structures using **Cn3D**

The remainder of the handout includes the introductory slides and the screen shots of the exercise demonstrated in Problem 1.

URL: http://www.ncbi.nlm.nih.gov/Class/minicourses/quickstructure.html

Course developed by: Dr. Medha Bhagwat (bhagwat@ncbi.nlm.nih.gov)

Slides:



http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

Conserved Domain

- recurring unit in molecular evolution, whose extents can be determined by sequence and structure analysis
- performs a particular function
- represented as a multiple local sequence alignment of proteins containing the domain

NCBI

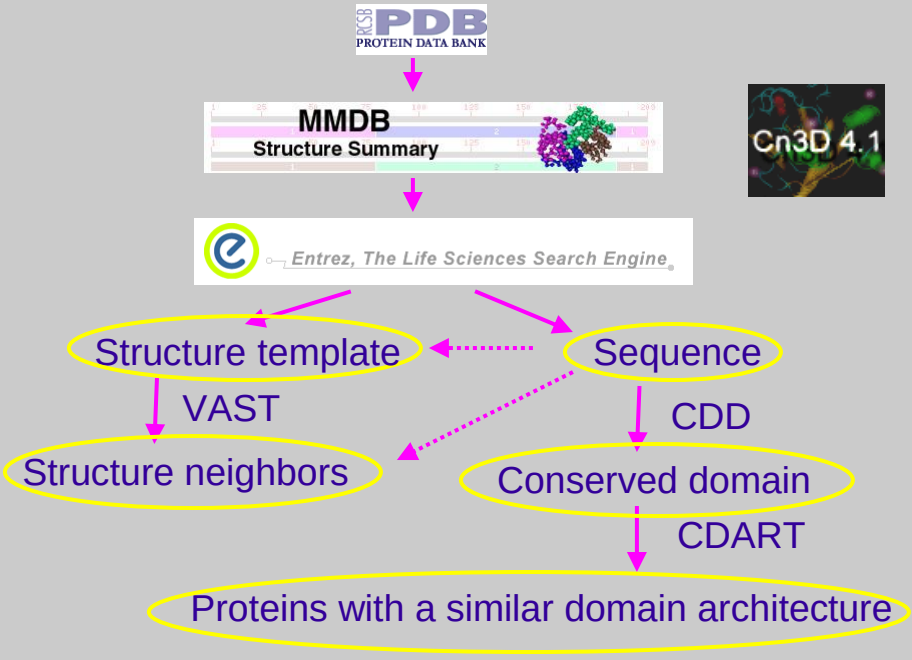
Conserved Domain Database

Pfam SMART COG KOG PRK

Curated CDs

- A position-specific scoring matrix (PSSM) is calculated
- CD-Search can be used to search against the PSSMs
- Manual curation of CDs has begun

NCBI



Problem 1

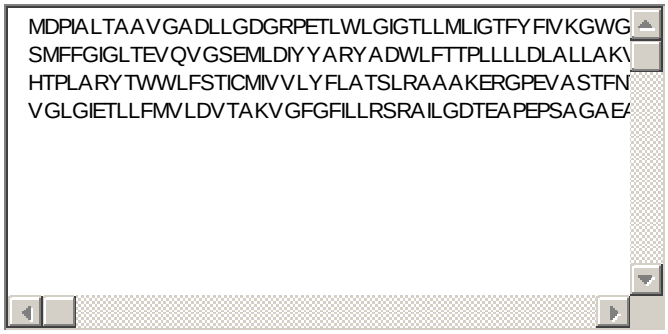
In this problem, we will follow these steps:

A. Identify conserved domain(s) present in a protein.
B. Search for other proteins containing similar domain(s).
C. Explore a 3D modeling template for the query sequence.
D. Find distant sequence homologs that may not be identified by BLAST.

NCBI's Conserved Domain Search allows you to match your protein sequence to a library of conserved protein domains, generate a multiple sequence alignment based on this match, and explore 3D modeling templates for your sequence. Click on the CDD link provided below,

CDD

Paste the following protein sequence in the CD-Search query box and run the search.



A. What is the domain present in this protein?
Obtain more information about the domain by searching in NCBI's Bookshelf

B. Go back to the CD-Search results page. Obtain a list of proteins with similar domain architecture by clicking on the "Search for similar domain architectures" button. To display the records, click on the link to the sequences and from there on the “Look up Sequences in Entrez”. Change the display from “Summary” to “FASTA”.

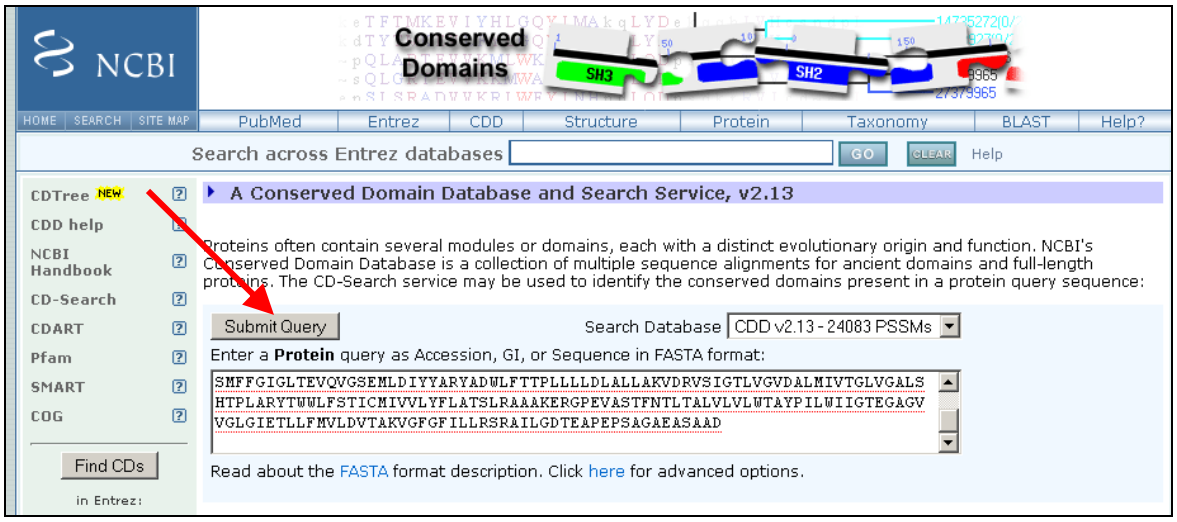
C. Go back to the CD-Search results page. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the graphic of the domain. Use the "Row Display" list box pull down menu to specify "up to 5" sequences and reformat sequence alignment. Extend the “Structure” display and invoke Cn3D with a display of a 3D modeling template and a multiple sequence alignment including your query sequence by pressing the "Show Structure" button.

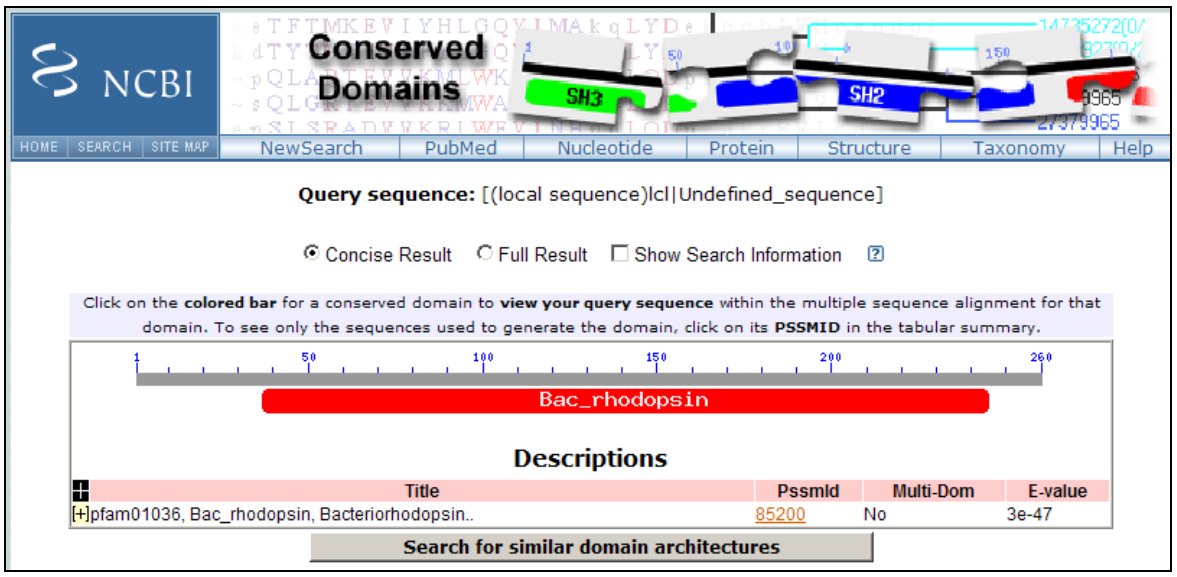
The structure of the *Halobacterium salinarum* bacteriorhodopsin mutant protein and its sequence alignment with our query protein are displayed. For a better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a bacterial rhodopsin signature (FMVLDVTAKVGF) where K is the retinal binding site. Identify these residues in the query protein and highlight the corresponding lysine residue in the halorhodopsin protein sequence.

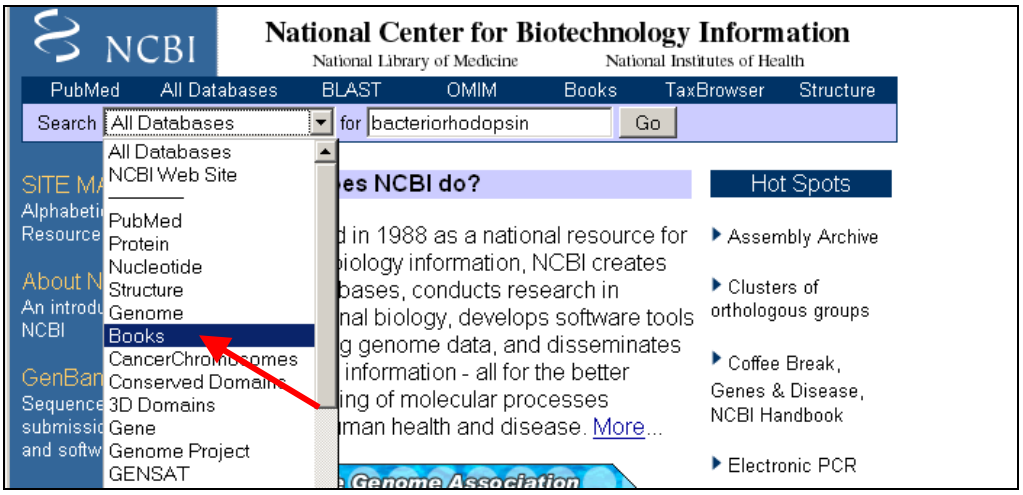
Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click the Protein Side chains on. Click on the “Done” button. To zoom in, press z on the keyboard. Identify the cofactor near the lysine residue.

D. To obtain the structural neighbors for the halorhodopsin protein, first click on the structure entry link, 1S52_B, on the CD-Browser page. Then click Links → Structure on the top right, then on 1S52 again in the Entrez Structure page, and finally on the chain A graphic. Select one or more of the check boxes next to the structure neighbors and view by clicking on the "View 3D Structure" button.

Screen images:





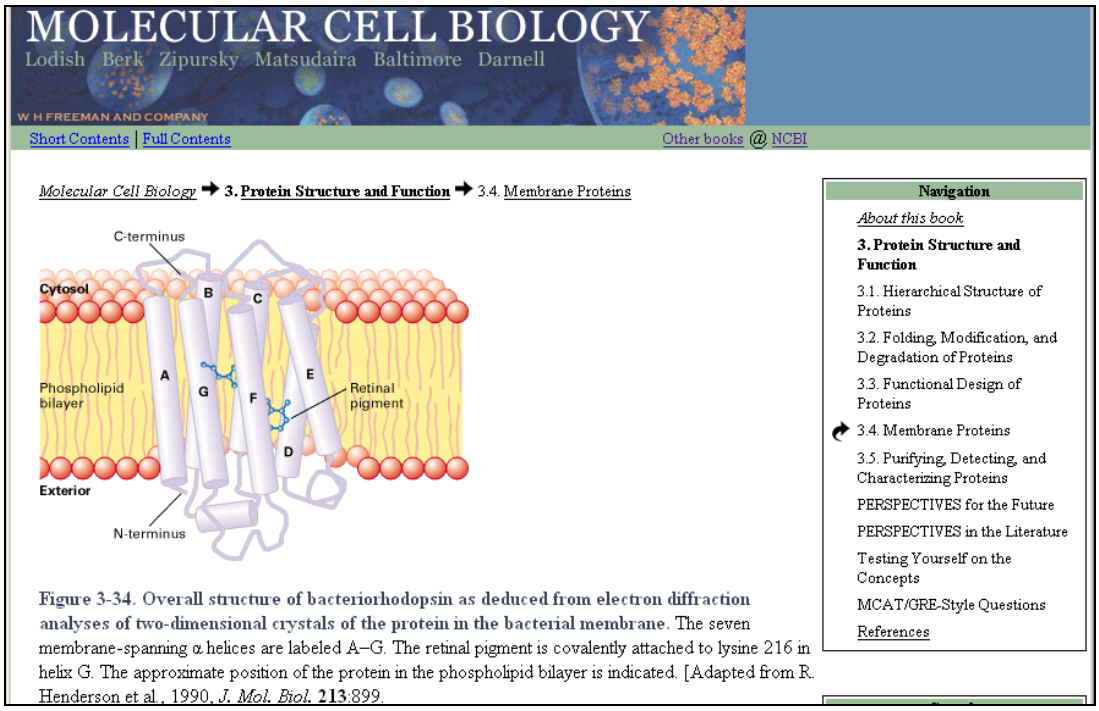


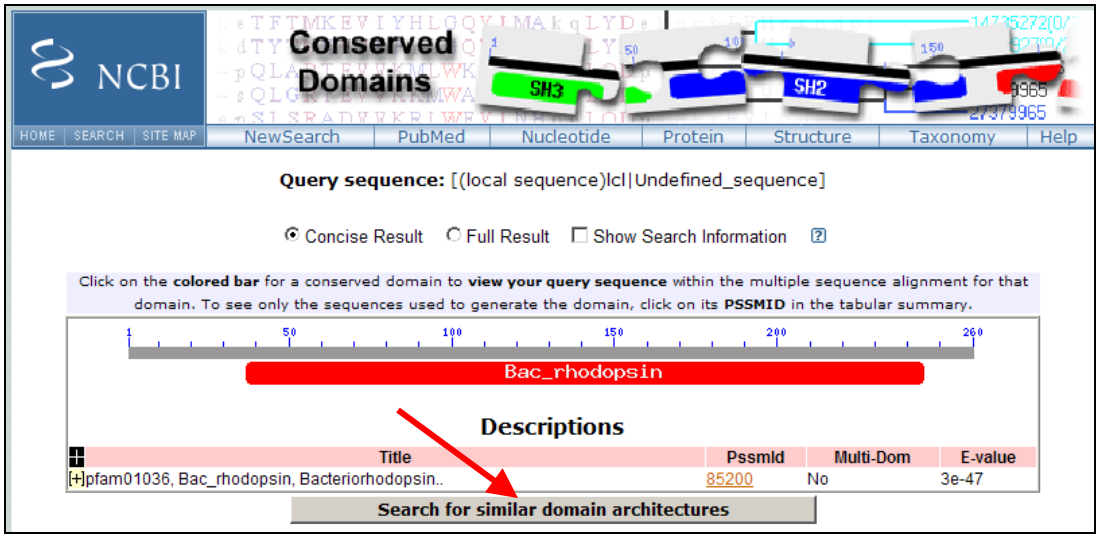
Many Integral Proteins Contain Multiple Transmembrane α Helices

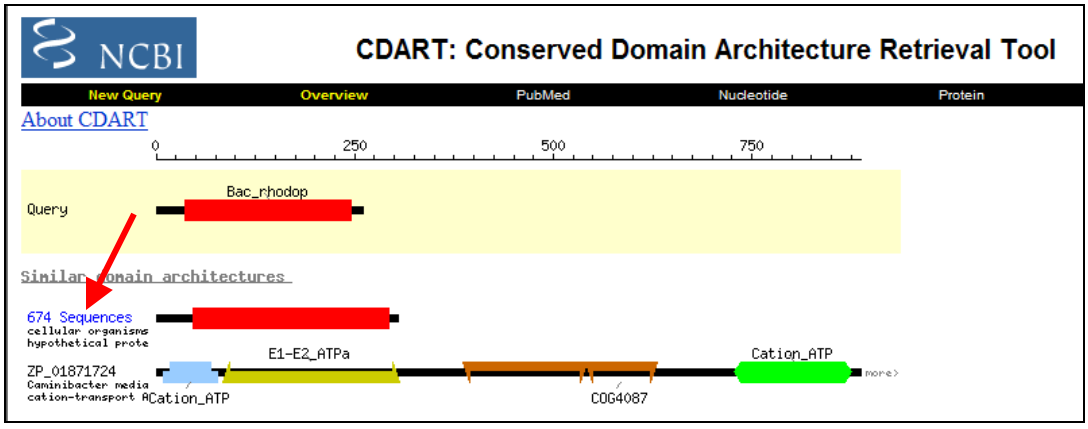
Although Figure 3-33 depicts glycophorin as a monomer with a single α helix spanning the bilayer, this protein is present in erythrocyte membranes as a dimer of two identical polypeptide chains. The two membrane-spanning α helices of glycophorin are thought to form a coiled-coil structure (see Figure 3-9a) stabilized by specific interactions between the amino acid side chains at the interface of the two helices. It is now known that many other transmembrane proteins contain two or more membrane-spanning α helices. For instance, the *bacterial photosynthetic reaction center (PRC)* comprises four subunits and several prosthetic groups, including four chlorophyll molecules. In this complex protein, three of the four subunits span the membrane; two of these subunits (L and M) each contain five membrane-spanning α helices (see Figure 16-40).

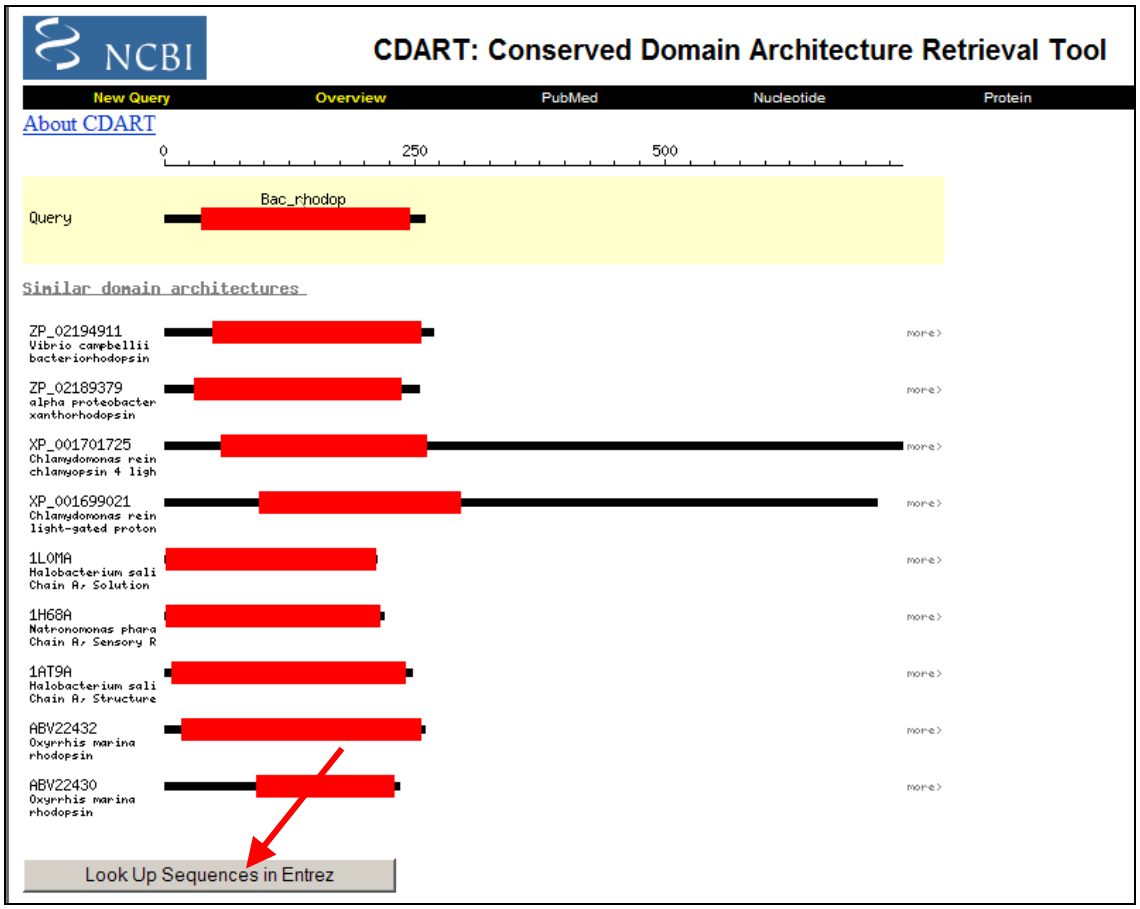
A large and important family of integral proteins is defined by the presence of seven membrane-spanning α helices. More than 150 such "seven-spanning" membrane proteins have been identified. This class of integral proteins is typified by *bacteriorhodopsin*, a protein found in a photosynthetic bacterium (Figure 3-34). Absorption of light by the retinal group attached to bacteriorhodopsin causes a conformational change in the protein that results in pumping of protons from the cytosol across the bacterial membrane to the extracellular space. The proton concentration gradient thus generated across the membrane is used to synthesize ATP, as discussed in Chapter 16. Both the overall arrangement of the seven α helices in bacteriorhodopsin and the identity of most of the amino acids can be resolved by computer analysis of micrographs of two-dimensional crystals of the membrane-embedded protein taken at various angles to the electron beam.

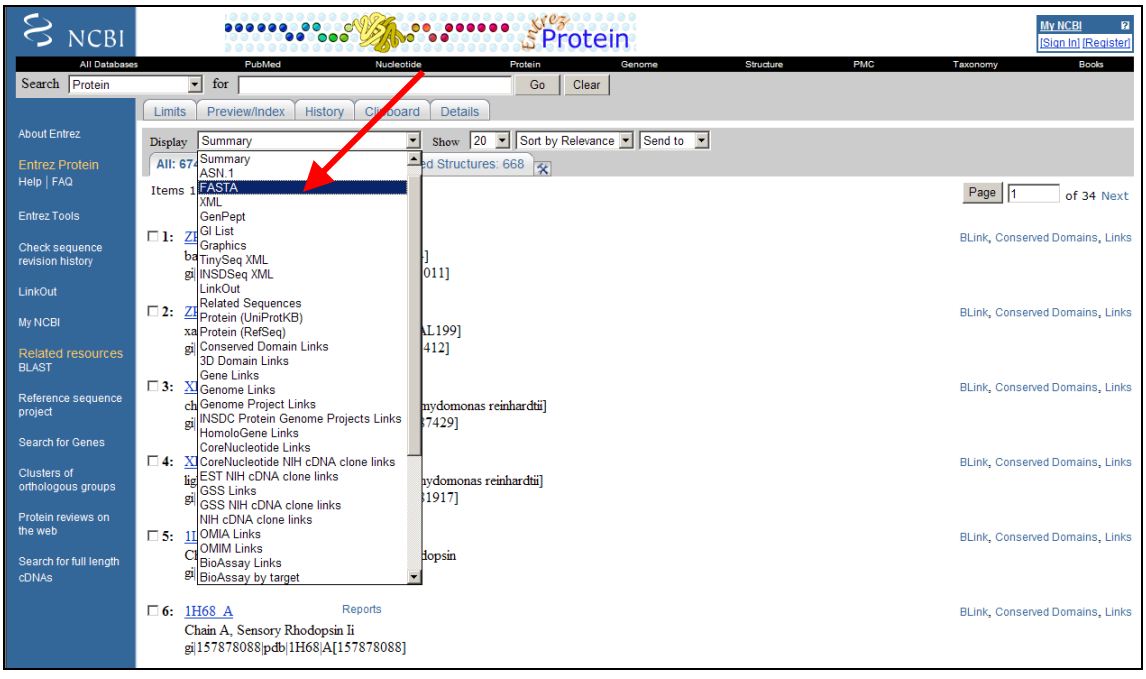
Other seven-spanning membrane proteins include the opsins (eye proteins that absorb light), cell-surface receptors for many hormones, and receptors for odorous molecules. Amino acid sequence analysis of these proteins has shown that no amino acids are found in the same position in all of them, and only a few residues are conserved in even a substantial number of them. Nonetheless, each of these proteins contains seven stretches of hydrophobic amino acids long enough (>22 amino acids) to span the phospholipid bilayer. Though direct evidence is lacking, it is thought that all of these proteins adopt a conformation in the membrane similar to that of bacteriorhodopsin. This is one of several examples of how investigators can predict the orientation of proteins in a membrane from the amino acid sequence alone. ↑ TOP

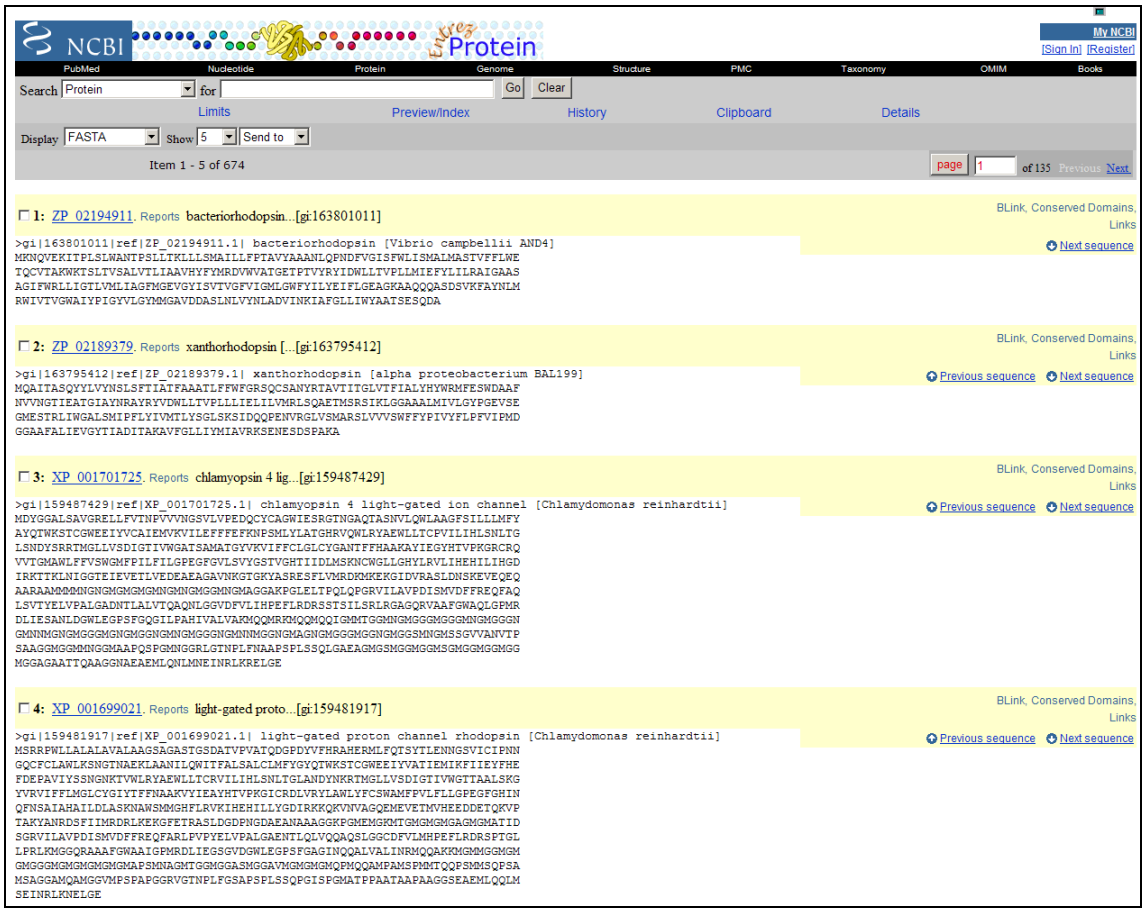

MOLECULAR CELL BIOLOGY
Lodish Berk Zipursky Matsudaira Baltimore Darnell
W H FREEMAN AND COMPANY
Short Contents | Full Contents
Other books @ NCBI
Molecular Cell Biology ➔ 3. Protein Structure and Function ➔ 3.4. Membrane Proteins
C-terminus
Cytosol
Phospholipid bilayer
Exterior
N-terminus
Retinal pigment
A
B
C
D
E
F
G
Navigation
About this book
3. Protein Structure and Function
3.1. Hierarchical Structure of Proteins
3.2. Folding, Modification, and Degradation of Proteins
3.3. Functional Design of Proteins
3.4. Membrane Proteins
3.5. Purifying, Detecting, and Characterizing Proteins
PERSPECTIVES for the Future
PERSPECTIVES in the Literature
Testing Yourself on the Concepts
MCAT/GRE-Style Questions
References
Figure 3-34. Overall structure of bacteriorhodopsin as deduced from electron diffraction analyses of two-dimensional crystals of the protein in the bacterial membrane. The seven membrane-spanning α helices are labeled A–G. The retinal pigment is covalently attached to lysine 216 in helix G. The approximate position of the protein in the phospholipid bilayer is indicated. [Adapted from R. Henderson et al., 1990, J. Mol. Biol. 213:899.

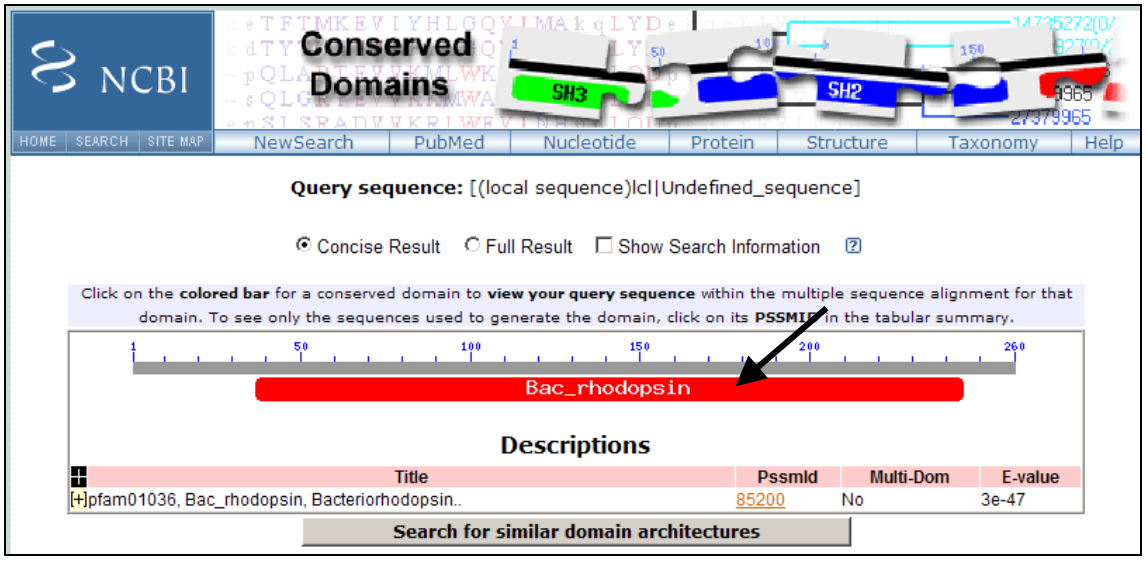

NCBI
Conserved Domains
SH3
SH2
HOME SEARCH SITE MAP
NewSearch PubMed Nucleotide Protein Structure Taxonomy Help
Query sequence: [(local sequence)lcl|Undefined_sequence]
Concise Result Full Result Show Search Information
Click on the colored bar for a conserved domain to view your query sequence within the multiple sequence alignment for that domain. To see only the sequences used to generate the domain, click on its PSSMID in the tabular summary.
1 50 100 150 200 260
Bac_rhodopsin
Descriptions
Title Pssmid Multi-Dom E-value
[+]pfam01036, Bac_rhodopsin, Bacteriorhodopsin.. 85200 No 3e-47
Search for similar domain architectures

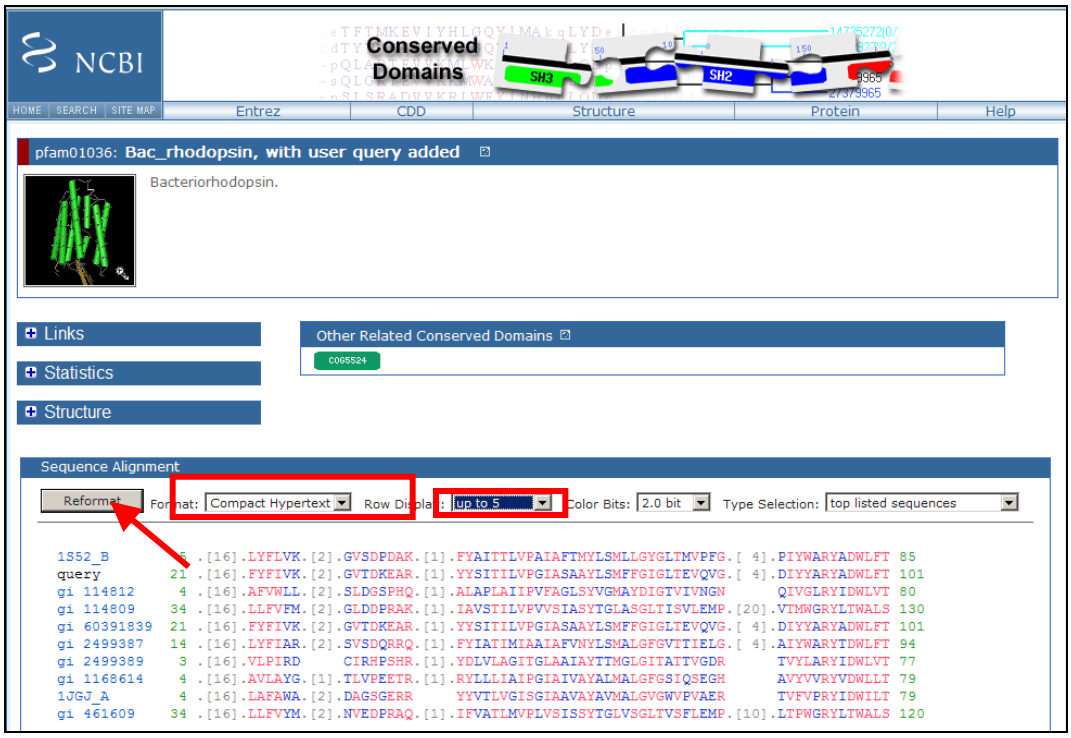

NCBI
CDART: Conserved Domain Architecture Retrieval Tool
New Query Overview PubMed Nucleotide Protein
About CDART
0 250 500 750
Query
Bac_rhodop
Similar domain architectures
674 Sequences
cellular organisms
hypothetical prote
ZP_01871724
Caminibacter media
cation-transport ACation_ATP
E1-E2_ATPa
Cation_ATP
COG4087
more

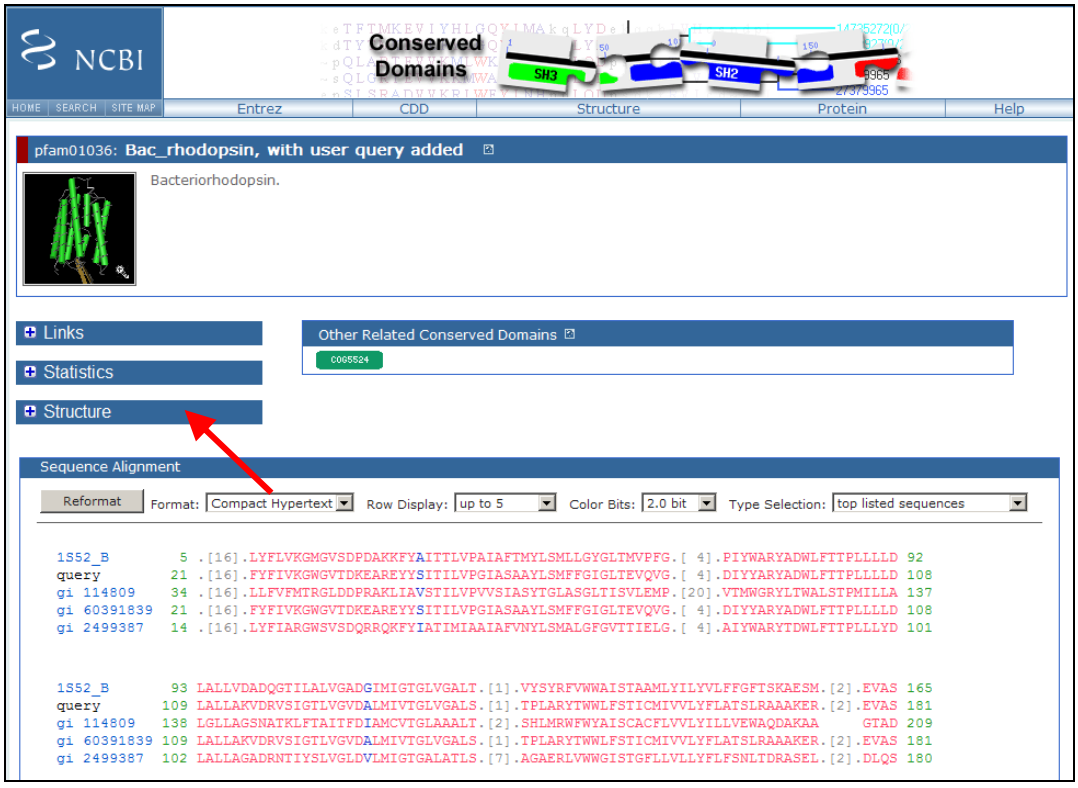

NCBI
CDART: Conserved Domain Architecture Retrieval Tool
New Query
Overview
PubMed
Nucleotide
Protein
About CDART
0
250
500
Query
Bac_rhodop
Similar domain architectures
ZP_02194911
Vibrio campbellii bacteriorhodopsin
ZP_02189379
alpha proteobacter xanthorhodopsin
XP_001701725
Chlamydomonas rein chlamyopsin 4 ligh
XP_001699021
Chlamydomonas rein light-gated proton
1L0MA
Halobacterium sali Chain A, Solution
1H68A
Natronomonas phara Chain A, Sensory R
1AT9A
Halobacterium sali Chain A, Structure
ABV22432
Oxyrrhis marina rhodopsin
ABV22430
Oxyrrhis marina rhodopsin
more>
Look Up Sequences in Entrez

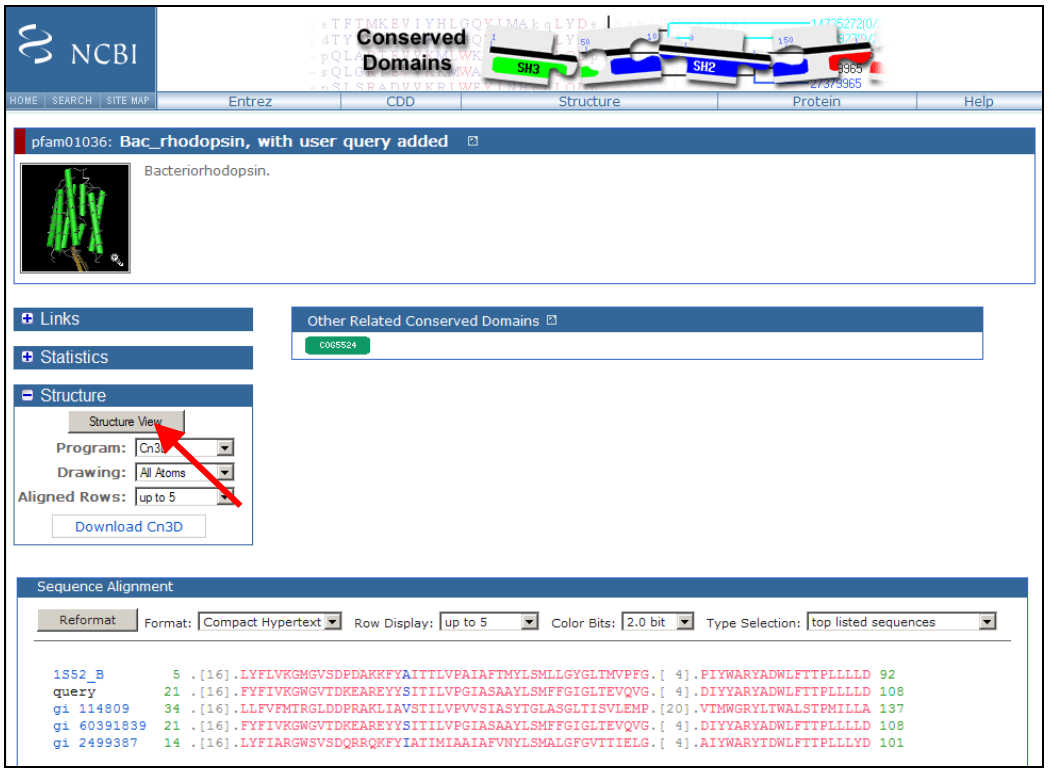

NCBI
Entrez Protein
My NCBI
[Sign In] [Register]
All Databases
PubMed
Nucleotide
Protein
Genome
Structure
PMC
Taxonomy
Books
Search Protein for
Go
Clear
Limits
Preview/Index
History
Clipboard
Details
About Entrez
Entrez Protein
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI
Related resources
BLAST
Reference sequence project
Search for Genes
Clusters of orthologous groups
Protein reviews on the web
Search for full length cDNAs
Display
Summary
Show
20
Sort by Relevance
Send to
Page
1
of 34
Next
Summary
ASN.1
FASTA
XML
GenPept
GI List
Graphics
TinySeq XML
INSDSeq XML
LinkOut
Related Sequences
Protein (UniProtKB)
Protein (RefSeq)
Conserved Domain Links
3D Domain Links
Gene Links
Genome Links
Genome Project Links
INSDC Protein Genome Projects Links
HomoloGene Links
CoreNucleotide Links
CoreNucleotide NIH cDNA clone links
EST NIH cDNA clone links
GSS Links
GSS NIH cDNA clone links
NIH cDNA clone links
OMIA Links
OMIM Links
BioAssay Links
BioAssay by target
BLink, Conserved Domains, Links
1H68_A
Reports
Chain A, Sensory Rhodopsin Ii
gi|157878088|pdb|1H68|A[157878088]

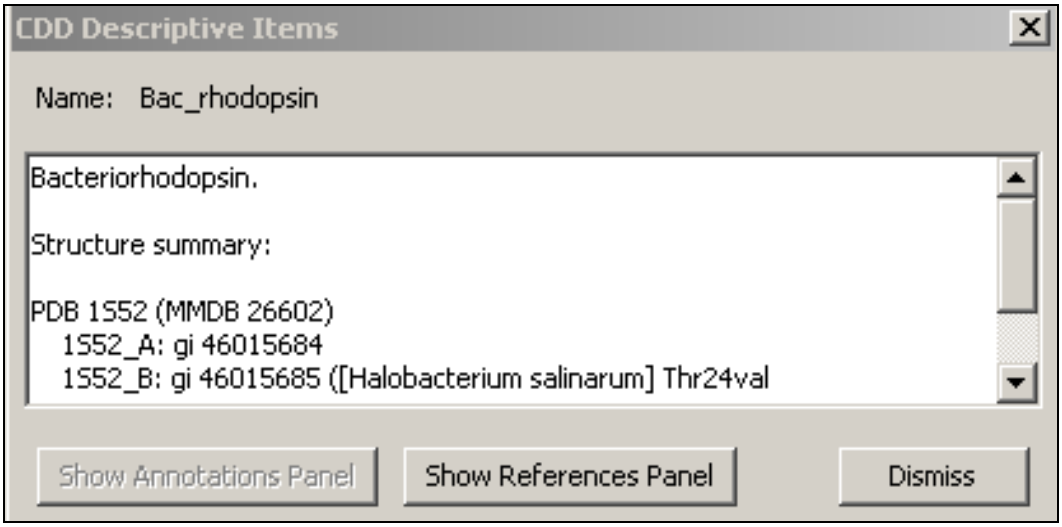

NCBI
Protein
My NCBI
[Sign In] [Register]
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Display FASTA Show 5 Send to
Item 1 - 5 of 674
page 1 of 135 Previous Next
1: ZP_02194911. Reports bacteriorhodopsin...[gi:163801011]
BLink, Conserved Domains, Links
>gi|163801011|ref|ZP_02194911.1| bacteriorhodopsin [Vibrio campbellii AND4]
MKNQVEKITPLSLWANTPSLITKLLLSMAILLFPTAVYAAANLQPNDFVGISFWLISMALMASTVFFLWE
TQCVTAKWKTSLITVSALVTLIAAVHYFYMRDVWVATGETPTVYRYIDWLLTVPLLMIEFYLILRAIGAAS
AGIFWRLLIGTLVMLIAGFMGEVGYISVTVGFVIGMLGWFYILYEIFLGEAGKAAQQQASDSVKFAYNLM
RWIVTVGWAIYPIGYVLGYMMGAVDDASLNLVYNLADVINKIAFGLLIWYAATSESQDA
2: ZP_02189379. Reports xanthorhodopsin [...[gi:163795412]
BLink, Conserved Domains, Links
>gi|163795412|ref|ZP_02189379.1| xanthorhodopsin [alpha proteobacterium BAL199]
Previous sequence Next sequence
MQAITASQYYLVYNSLSFTIATFAAATLFFWFGRSQCSANYRTAVTITGLVTFIALYHYWRMFESWDAAF
NVVNGTIEATGIAYNRAYRYVDWLLTVPLLLIELILVMRLSQAETMSRSIKLGGAAALMIVLGYPGEVSE
GMESTRLIWGALSMIPFLYIVMTLYSGLSKSIDQQPENVRGLVSMARSLVVVSWFFYPIVYFLPFVIPMD
GGAAFALIEVGYTIADITAKAVFGLLIYMIAVRKSENESDSPAKA
3: XP_001701725. Reports chlamyopsin 4 lig...[gi:159487429]
BLink, Conserved Domains, Links
>gi|159487429|ref|XP_001701725.1| chlamyopsin 4 light-gated ion channel [Chlamydomonas reinhardtii]
Previous sequence Next sequence
MDYGGALSAVGRELLFVTNPVVVNGSVLVPEDQCYCAGWIESRGTNGAQTASNVLQWLAAGFSILLLMFY
AYQTWKSTCGWEEIYVCAIEMVKVILEFFFEFKNPSMLYLATGHRVQWLRYAEWLLTCPVILIHLSNLTG
LSNDYSRRTMGLLVSDIGTIVWGATSAMATGYVKVIFFCLGLCYGANTFFHAAKAYIEGYHTVPKGRCRQ
VVTGMAWLFFVSWGMFPILFILGPEGFGVLSVYGSTVGHTIIDLMSKNCWGLLGHYLRVLIHEHILIHGD
IRKTTKLNIGGTEIEVETLVEDEAEAGAVNKGTGKYASRESFLVMRDKMKEKGIDVRASLDNSKEVEQEQ
AARAAMMMNGNGMGMGMGMNGMGGMNGMAGGAKPGLELTPQLQPGRVILAVPDISMVDFFREQFAQ
LSVTYELVPALGADNTLALVTQAQNLGGVDFVLIHPEFLRDRSSTSILSRLRGAGQRVAAFGWAQLGPMR
DLIESANLDGWLEGPSFGQGILPAHIVALVAKMQQMRKMQQMQQIGMMTGGMNGMGGGMGGGMNGMGGGN
GMNNMGNGMGGGMGNGMGGNGMNGMGGGNGMNNMGGNGMAGNGMGGGMGGNGMGGSMNGMSSGVVANVTP
SAAGGMGGMMNGGMAAPQSPGMNGGRLGTNPLFNAAPSPLSSQLGAEAGMGSMGGMGGMSGMGGMGGMGG
MGGAGAATTQAAGGNAEAEMLQNLMNEINRLKRELGE
4: XP_001699021. Reports light-gated proto...[gi:159481917]
BLink, Conserved Domains, Links
>gi|159481917|ref|XP_001699021.1| light-gated proton channel rhodopsin [Chlamydomonas reinhardtii]
Previous sequence Next sequence
MSRRPWLLALALAVALAAGSAGASTGSDATVPVATQDGPDYVFHRAHERMLFQTSYTLENNGSVICIPNN
GQCFCLAWLKSNGTNAEKLAANILQWITFALSALCLMFYGYQTWKSTCGWEEIYVATIEMIKFIIEYFHE
FDEPAVIYSSNGNKTVWLRYAEWLLTCRVILIHLSNLTGLANDYNKRTMGLLVSDIGTIVWGTTAALSKG
YVRVIFFLMGLCYGIYTFFNAAKVYIEAYHTVPKGICRDLVRYLAWLYFCSWAMFPVLFLLGPEGFGHIN
QFNSAIAHAILDLASKNAWSMMGHFLRVKIHEHILLYGDIRKKQKVNVAGQEMEVETMVHEEDDETQKVP
TAKYANRDSFIIMRDRLKEKGFETRASLDGDPNGDAEANAAAGGKPGMEMGKMTGMGMGMGAGMGMATID
SGRVILAVPDISMVDFFREQFARLPVPYELVPALGAENTLQLVQQAQSLGGCDFVLMHPEFLRDRSPTGL
LPRLKMGGQRAAAFGWAAIGPMRDLIEGSGVDGWLEGPSFGAGINQQALVALINRMQQAKKMGMMGGMGM
GMGGGMGMGMGMGMAPSMNAGMTGGMGGASMGGAVMGMGMGMQPMQQAMPAMSPMMTQQPSMMSQPSA
MSAGGAMQAMGGVMPSPAPGGRVGTNPLFGSAPSPLSSQPGISPGMATPPAATAAPAAGGSEAEMLQQLM
SEINRLKNELGE

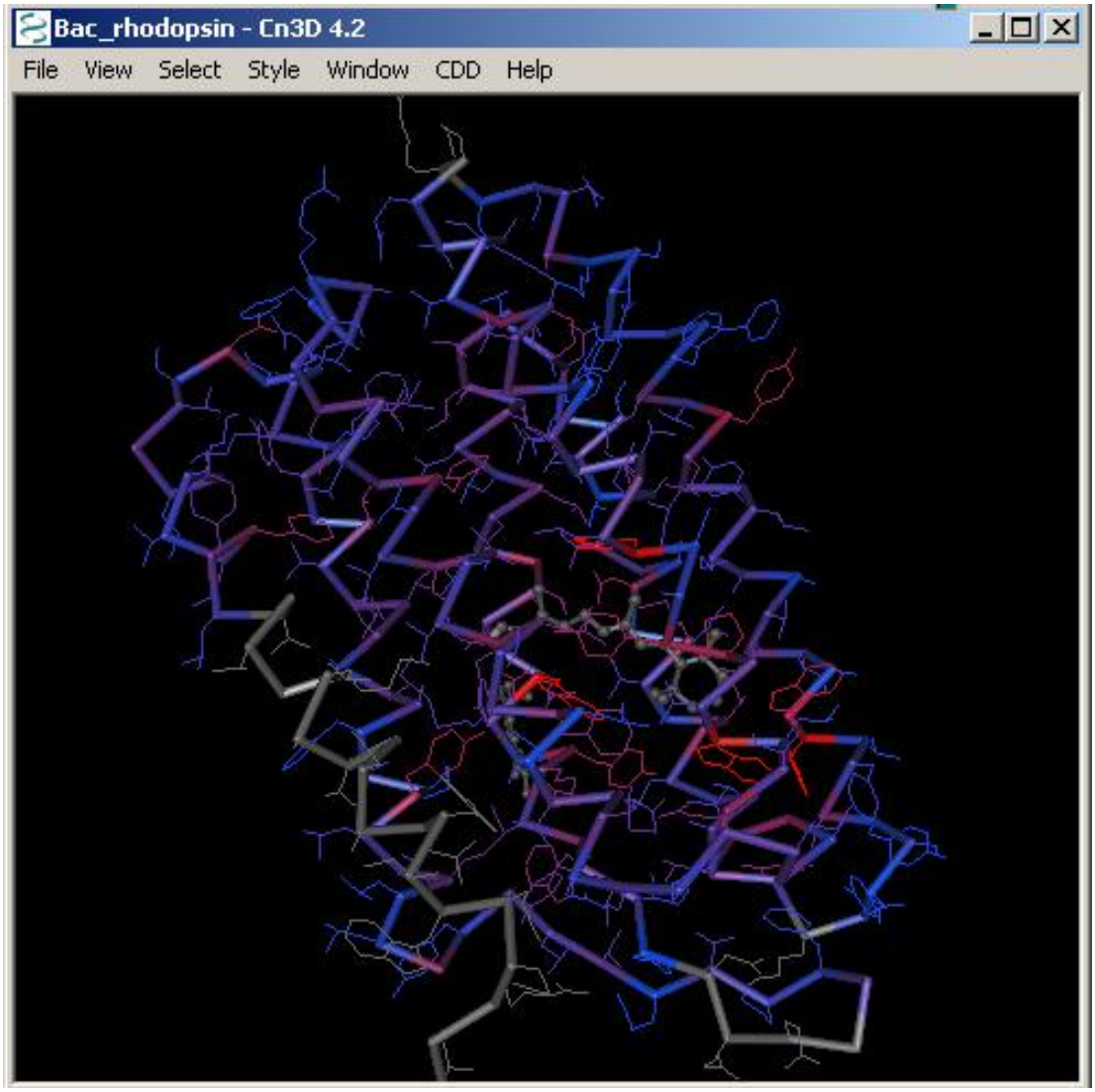

NCBI
Conserved Domains
SH3
SH2
HOME SEARCH SITE MAP
NewSearch PubMed Nucleotide Protein Structure Taxonomy Help
Query sequence: [(local sequence)lcl|Undefined_sequence]
Concise Result Full Result Show Search Information
Click on the colored bar for a conserved domain to view your query sequence within the multiple sequence alignment for that domain. To see only the sequences used to generate the domain, click on its PSSMID in the tabular summary.
1 50 100 150 200 260
Bac_rhodopsin
Descriptions
Title PssmId Multi-Dom E-value
[+]pfam01036, Bac_rhodopsin, Bacteriorhodopsin.. 85200 No 3e-47
Search for similar domain architectures

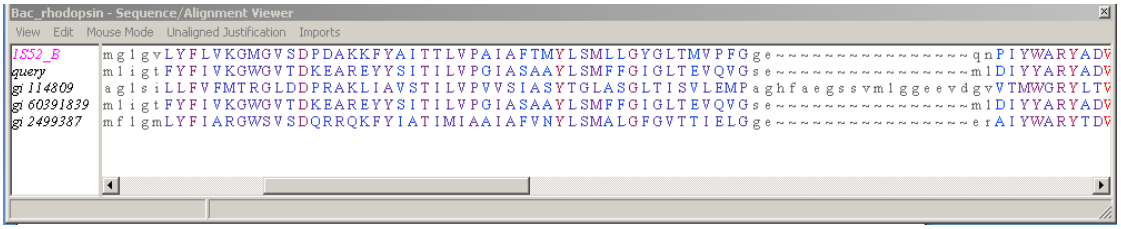

NCBI
Conserved Domains
SH3
SH2
HOME
SEARCH
SITE MAP
Entrez
CDD
Structure
Protein
Help
pfam01036: Bac_rhodopsin, with user query added
Bacteriorhodopsin.
Links
Statistics
Structure
Other Related Conserved Domains
Sequence Alignment
Reformat
Format: Compact Hypertext
Row Display: up to 5
Color Bits: 2.0 bit
Type Selection: top listed sequences
1S52_B 5 .[16].LYFLVK.[2].GVSDPDAK.[1].FYAITTLVPAIAFTMYLSMLLGYGLTMVPFG.[4].PIYWARYADWLFT 85
query 21 .[16].FYFIVK.[2].GVTDKEAR.[1].YYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFT 101
gi 114812 4 .[16].AFVWLL.[2].SLDGSPHQ.[1].ALAPLAIIPVFAGLSYVGMAYDIGTVIVNGN QIVGLRYIDWLVT 80
gi 114809 34 .[16].LLFVFM.[2].GLDDPRAK.[1].IAVSTILVPVVSIASYTGLASGLTISVLEMP.[20].VTMWGRYLTWALS 130
gi 60391839 21 .[16].FYFIVK.[2].GVTDKEAR.[1].YYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFT 101
gi 2499387 14 .[16].LYFIAR.[2].SVSDQRRQ.[1].FYIATIMIAAIAFVNYLSMALGFGVTTIELG.[4].AIYWARYTDWLFT 94
gi 2499389 3 .[16].VLPIRD CIRHPSHR.[1].YDLVLAGITGLAAIAYTTMGLGITATTVGDR TVYLARYIDWLVT 77
gi 1168614 4 .[16].AVLAYG.[1].TLVPEETR.[1].RYLLLIAIPGIAIVAYALMALGFGSIQSEGH AVYVVRYVDWLLT 79
1JGJ_A 4 .[16].LAFAWA.[2].DAGSGERR YYVTLVGISGIAAVAYAVMALGVGWVPVAER TVFVPRYIDWILT 79
gi 461609 34 .[16].LLFVYM.[2].NVEDPRAQ.[1].IFVATLMVPLVSISSYTGLVSGLTVSFLEMP.[10].LTPWGRYLTWALS 120

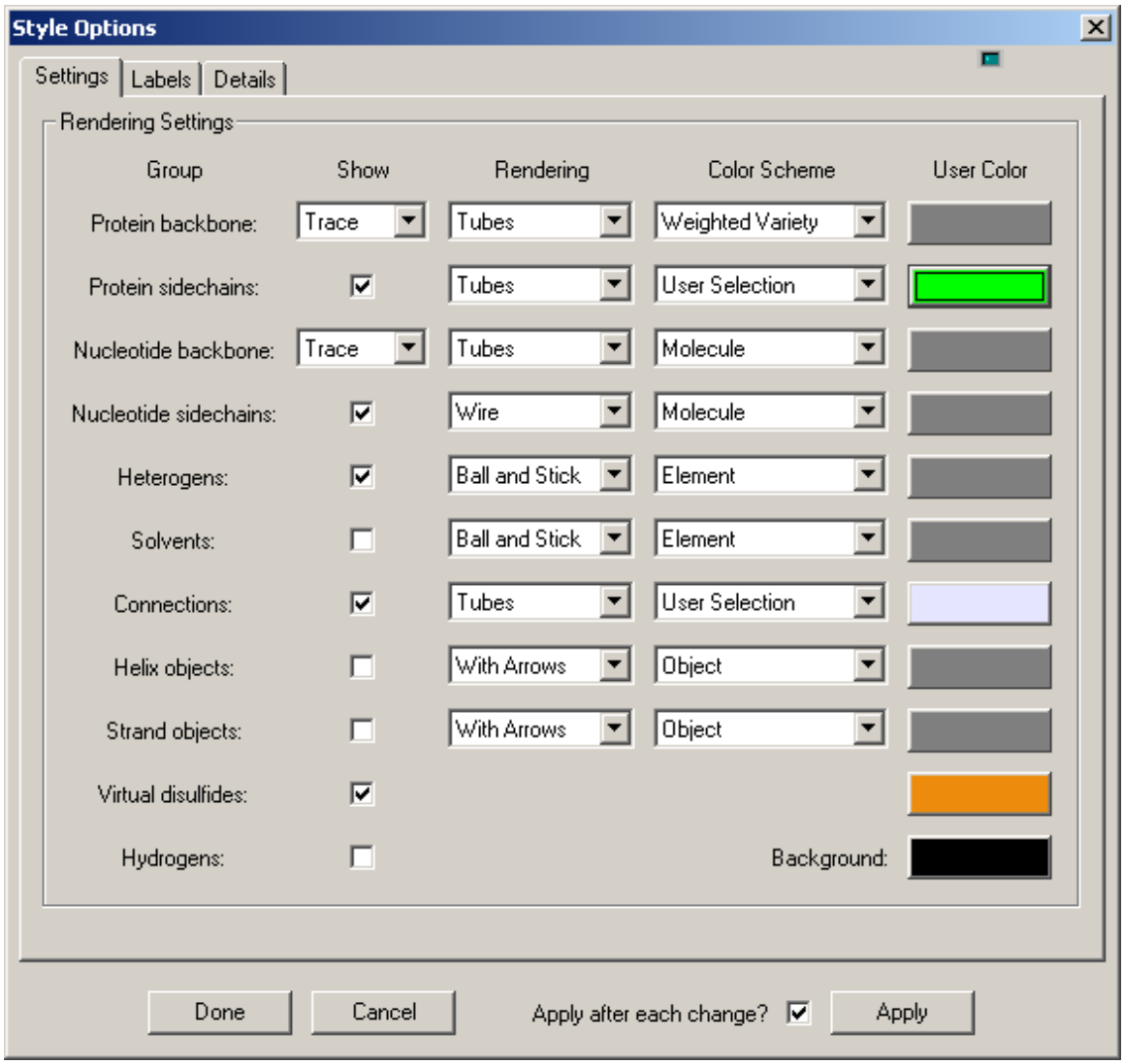

NCBI
Conserved Domains
SH3
SH2
HOME
SEARCH
SITE MAP
Entrez
CDD
Structure
Protein
Help
pfam01036: Bac_rhodopsin, with user query added
Bacteriorhodopsin.
Links
Statistics
Structure
Other Related Conserved Domains
Sequence Alignment
Reformat
Format: Compact Hypertext
Row Display: up to 5
Color Bits: 2.0 bit
Type Selection: top listed sequences
1S52_B 5 .[16].LYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVPFG.[4].PIYWARYADWLFTTPLLLLD 92
query 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLD 108
gi 114809 34 .[16].LLFVFMTRGLDDPRAKLIAVSTILVPVVSIASYTGLASGLTISVLEMP.[20].VTMWGRYLTWALSTPMILLA 137
gi 60391839 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLD 108
gi 2499387 14 .[16].LYFIARGWSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[4].AIYWARYTDWLFTTPLLLYD 101
1S52_B 93 LALLVDADQGTILALVGADGIMIGTGLVGALT.[1].VYSYRFVWWAISTAAMLYILYVLFFGFTSKAESM.[2].EVAS 165
query 109 LALLAKVDRVSIGTLVGVDALMIVTGLVGALS.[1].TPLARYTWWLFSTICMIVVLYFLATSLRAAAKER.[2].EVAS 181
gi 114809 138 LGLLAGSNATKLFTAITFDIAMCVTGLAAALT.[2].SHLMRWFWYAISCACFLVVLYILLVEWAQDAKAA GTAD 209
gi 60391839 109 LALLAKVDRVSIGTLVGVDALMIVTGLVGALS.[1].TPLARYTWWLFSTICMIVVLYFLATSLRAAAKER.[2].EVAS 181
gi 2499387 102 LALLAGADRNTIYSLVGLDVLMIGTGALATLS.[7].AGAERLVWWGISTGFLLVLLYFLFSNLTDRASEL.[2].DLQS 180

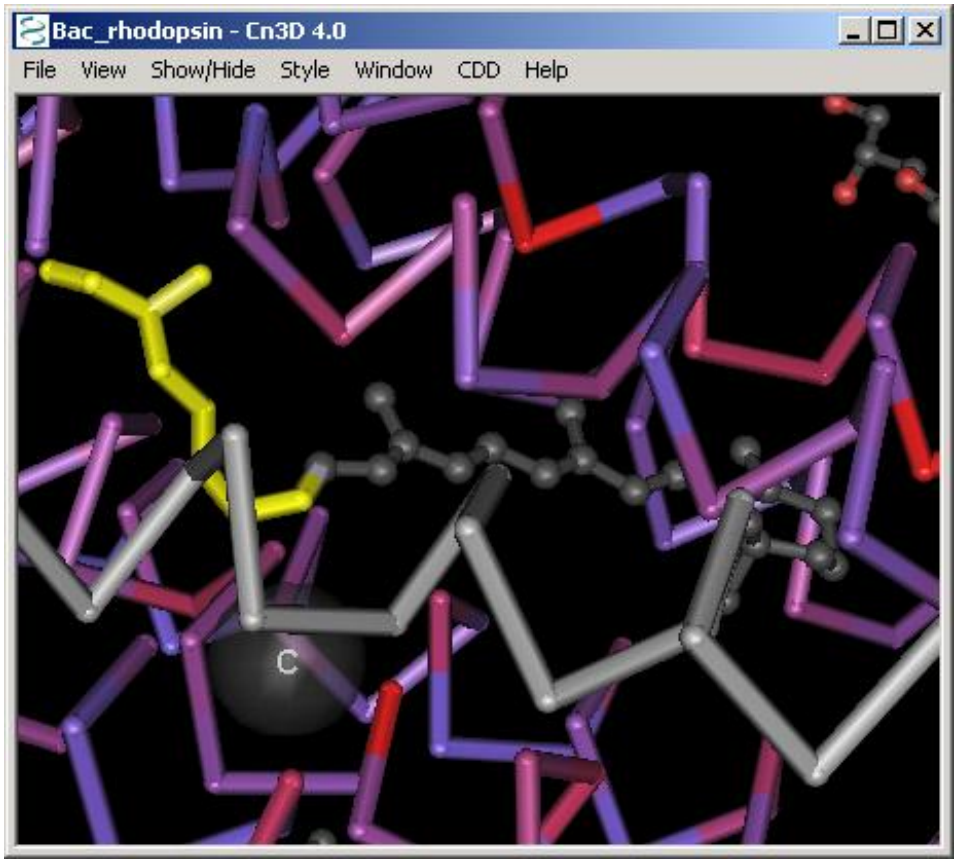

NCBI
Conserved Domains
SH3
SH2
HOME
SEARCH
SITE MAP
Entrez
CDD
Structure
Protein
Help
pfam01036: Bac_rhodopsin, with user query added
Bacteriorhodopsin.
Links
Statistics
Structure
Structure View
Program:
Drawing:
All Atoms
Aligned Rows:
up to 5
Download Cn3D
Other Related Conserved Domains
Sequence Alignment
Reformat
Format:
Compact Hypertext
Row Display:
up to 5
Color Bits:
2.0 bit
Type Selection:
top listed sequences
1S52_B 5 .[16].LYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVPFG.[4].PIYWARYADWLFTTPLLLLD 92
query 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLD 108
gi 114809 34 .[16].LLFVFMTRGLDDPRAKLIAVSTILVPVVSIASYTGLASGLTISVLEMP.[20].VTMWGRYLTWALSTPMILLA 137
gi 60391839 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLD 108
gi 2499387 14 .[16].LYFIARGWSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[4].AIYWARYTDWLFTTPLLLYD 101

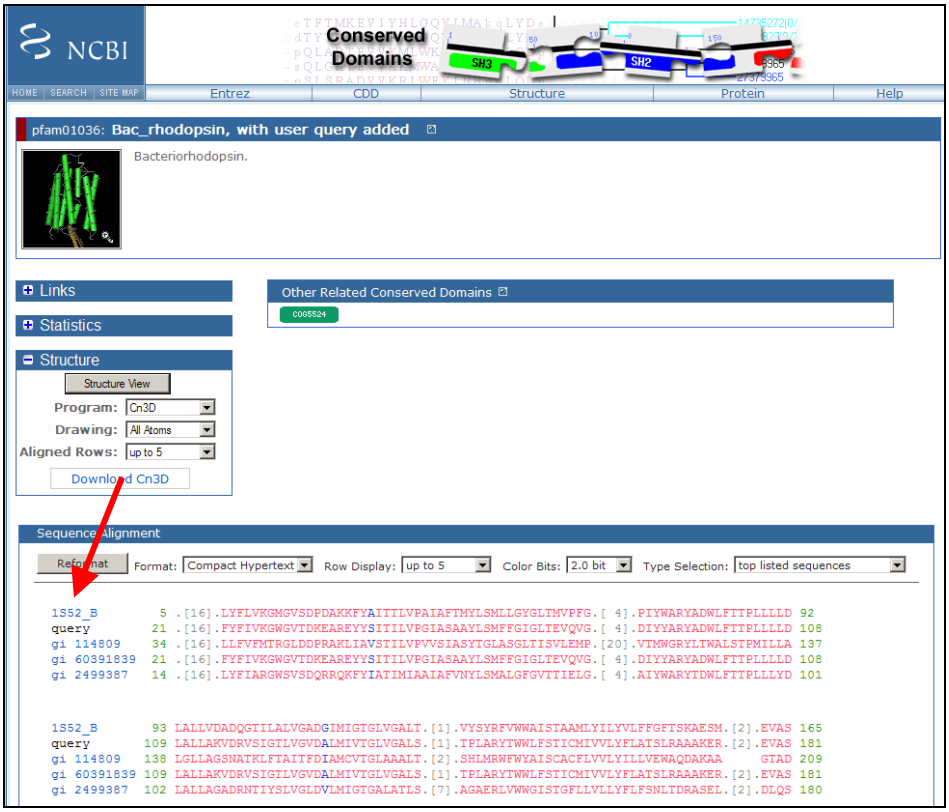

CDD Descriptive Items
Name: Bac_rhodopsin
Bacteriorhodopsin.
Structure summary:
PDB 1S52 (MMDB 26602)
1S52_A: gi 46015684
1S52_B: gi 46015685 ([Halobacterium salinarum] Thr24val
Show Annotations Panel
Show References Panel
Dismiss

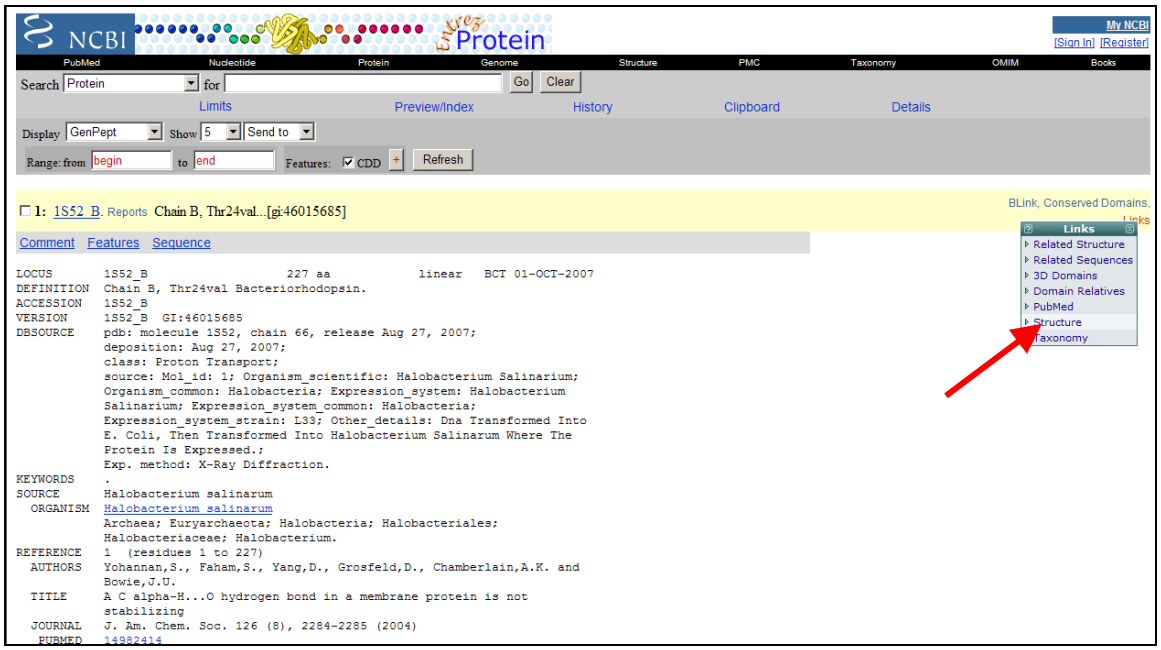

Bac_rhodopsin - Cn3D 4.2
File View Select Style Window CDD Help

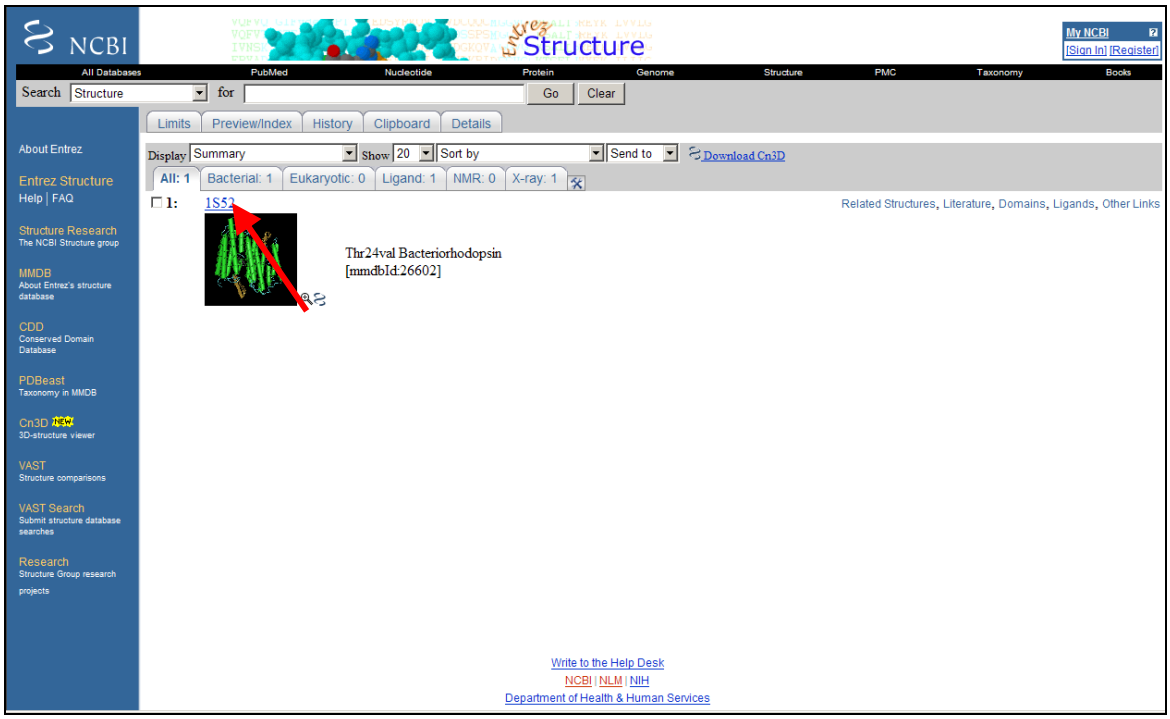

Bac_rhodopsin - Sequence/Alignment Viewer
View Edit Mouse Mode Unaligned Justification Imports
1S52_B
query
gi 114809
gi 60391839
gi 2499387

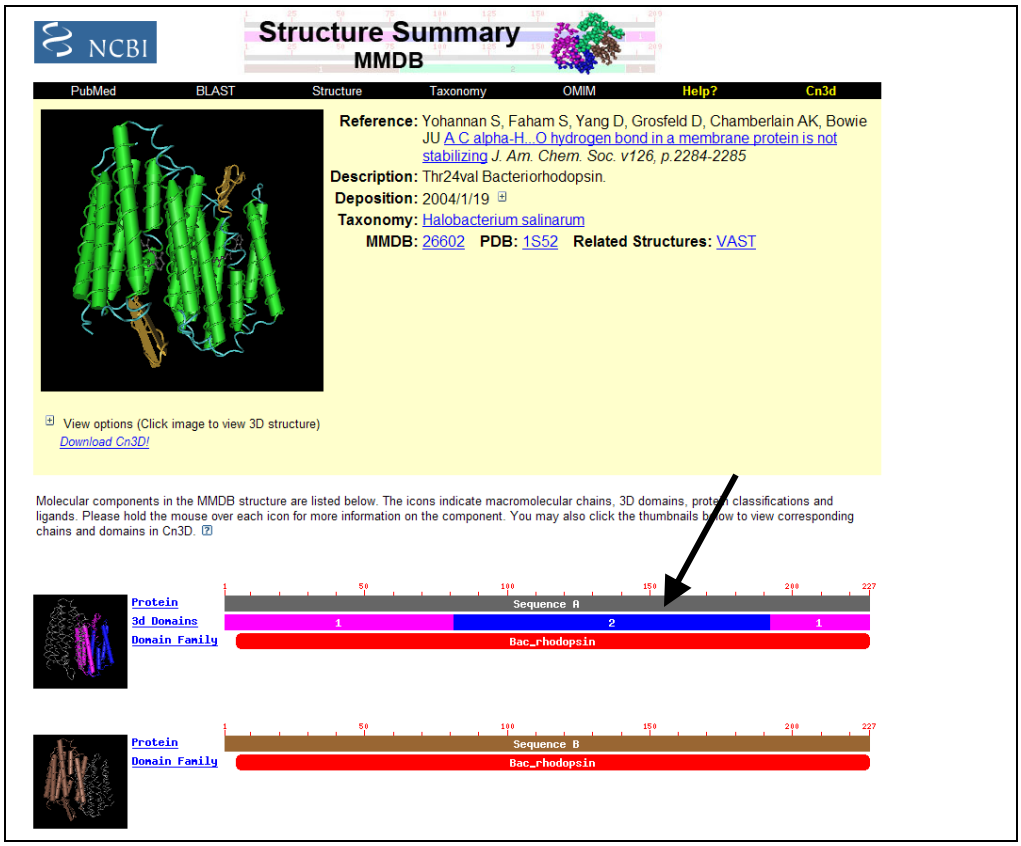

Style Options
Settings
Labels
Details
Rendering Settings
Group
Show
Rendering
Color Scheme
User Color
Protein backbone:
Trace
Tubes
Weighted Variety
Protein sidechains:
Tubes
User Selection
Nucleotide backbone:
Trace
Tubes
Molecule
Nucleotide sidechains:
Wire
Molecule
Heterogens:
Ball and Stick
Element
Solvents:
Ball and Stick
Element
Connections:
Tubes
User Selection
Helix objects:
With Arrows
Object
Strand objects:
With Arrows
Object
Virtual disulfides:
Hydrogens:
Background:
Done
Cancel
Apply after each change?
Apply

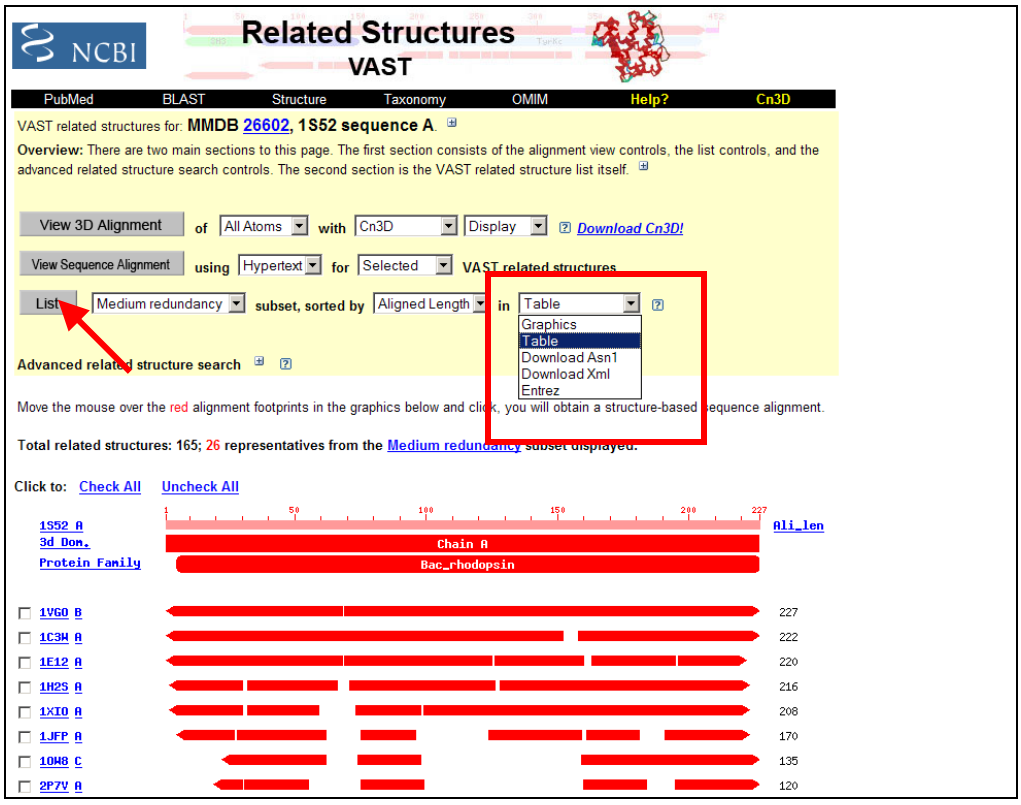

Bac_rhodopsin - Cn3D 4.0
File View Show/Hide Style Window CDD Help
C

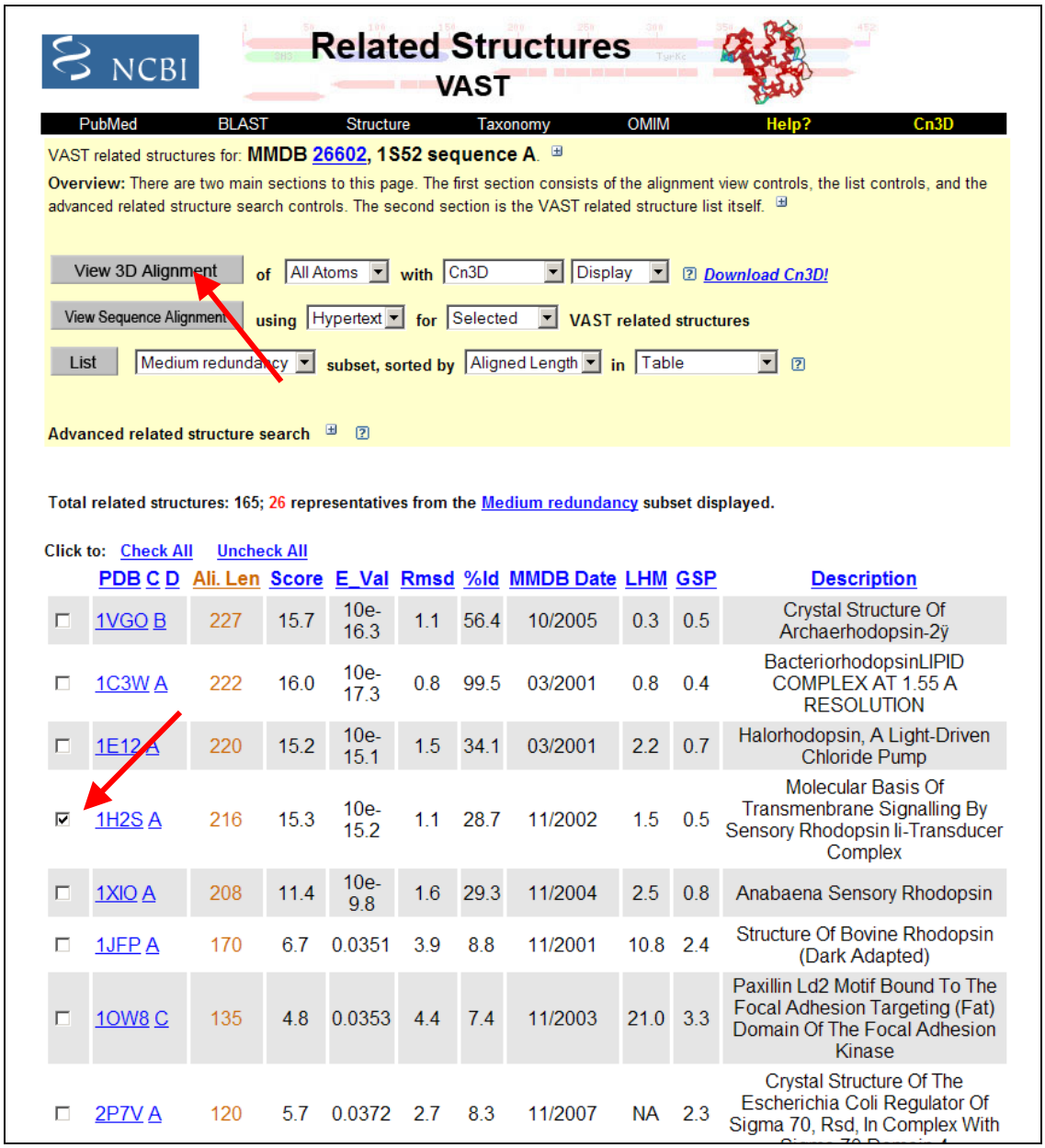

NCBI
Conserved Domains
SH3
SH2
HOME SEARCH SITE MAP
Entrez CDD Structure Protein Help
pfam01036: Bac_rhodopsin, with user query added
Bacteriorhodopsin.
Links
Statistics
Structure
Structure View
Program: Cn3D
Drawing: All Atoms
Aligned Rows: up to 5
Download Cn3D
Other Related Conserved Domains
Sequence Alignment
Reformat
Format: Compact Hypertext
Row Display: up to 5
Color Bits: 2.0 bit
Type Selection: top listed sequences
1S52_B 5 .[16].LYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVPFG.[4].PIYWARYADWLFTTPLLLLD 92
query 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLD 108
gi 114809 34 .[16].LLFVFMTRGLDDPRAKLIAVSTILVPVVSIASYTGLASGLTISVLEMP.[20].VTMWGRYLTWALSTPMILLA 137
gi 60391839 21 .[16].FYFIVKGWGVTDKEAREYYSITILVPGIASAAYLSMFFGIGLTEVQVG.[4].DIYYARYADWLFTTPLLLLD 108
gi 2499387 14 .[16].LYFIARGWSVSDQRRQKFYIATIMIAAIAFVNYLSMALGFGVTTIELG.[4].AIYWARYTDWLFTTPLLLYD 101
1S52_B 93 LALLVDADQGTILALVGADGIMIGTGLVGALT.[1].VYSYRFVWWAISTAAMLYILYVLFFGFTSKAESM.[2].EVAS 165
query 109 LALLAKVDRVSIGTLVGVDALMIVTGLVGALS.[1].TPLARYTWWLFSTICMIVVLYFLATSLRAAAKER.[2].EVAS 181
gi 114809 138 LGLLAGSNATKLFTAITFDIAMCVTGLAAALT.[2].SHLMRWFWYAISCACFLVVLYILLVEWAQDAKAA GTAD 209
gi 60391839 109 LALLAKVDRVSIGTLVGVDALMIVTGLVGALS.[1].TPLARYTWWLFSTICMIVVLYFLATSLRAAAKER.[2].EVAS 181
gi 2499387 102 LALLAGADRNTIYSLVGLDVLMIGTGALATLS.[7].AGAERLVWWGISTGFLLVLLYFLFSNLTDRASEL.[2].DLQS 180


NCBI
Protein
My NCBI
[Sign In] [Register]
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Display GenPept Show 5 Send to
Range: from begin to end Features: CDD Refresh
1: 1S52_B. Reports Chain B, Thr24val...[gi:46015685]
BLink, Conserved Domains, Links
Comment Features Sequence
LOCUS 1S52_B 227 aa linear BCT 01-OCT-2007
DEFINITION Chain B, Thr24val Bacteriorhodopsin.
ACCESSION 1S52_B
VERSION 1S52_B GI:46015685
DBSOURCE pdb: molecule 1S52, chain 66, release Aug 27, 2007;
deposition: Aug 27, 2007;
class: Proton Transport;
source: Mol_id: 1; Organism_scientific: Halobacterium Salinarium;
Organism_common: Halobacteria; Expression_system: Halobacterium
Salinarium; Expression_system_common: Halobacteria;
Expression_system_strain: L33; Other_details: Dna Transformed Into
E. Coli, Then Transformed Into Halobacterium Salinarum Where The
Protein Is Expressed.;
Exp. method: X-Ray Diffraction.
KEYWORDS .
SOURCE Halobacterium salinarum
ORGANISM Halobacterium salinarum
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halobacterium.
REFERENCE 1 (residues 1 to 227)
AUTHORS Yohannan,S., Faham,S., Yang,D., Grosfeld,D., Chamberlain,A.K. and
Bowie,J.U.
TITLE A C alpha-H...O hydrogen bond in a membrane protein is not
stabilizing
JOURNAL J. Am. Chem. Soc. 126 (8), 2284-2285 (2004)
PUBMED 14982414
Links
Related Structure
Related Sequences
3D Domains
Domain Relatives
PubMed
Structure
Taxonomy


NCBI
Structure
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search Structure for Go Clear
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Structure
Help | FAQ
Structure Research
The NCBI Structure group
MMDB
About Entrez's structure database
CDD
Conserved Domain Database
PDBeast
Taxonomy in MMDB
Cn3D NEW
3D-structure viewer
VAST
Structure comparisons
VAST Search
Submit structure database searches
Research
Structure Group research projects
Display Summary Show 20 Sort by Send to Download Cn3D
All: 1 Bacterial: 1 Eukaryotic: 0 Ligand: 1 NMR: 0 X-ray: 1
1: 1S52
Related Structures, Literature, Domains, Ligands, Other Links
Thr24val Bacteriorhodopsin
[mmdbId:26602]
Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services


NCBI
Structure Summary
MMDB
PubMed BLAST Structure Taxonomy OMIM Help? Cn3d
Reference: Yohannan S, Faham S, Yang D, Grosfeld D, Chamberlain AK, Bowie JU A C alpha-H...O hydrogen bond in a membrane protein is not stabilizing J. Am. Chem. Soc. v126, p.2284-2285
Description: Thr24val Bacteriorhodopsin.
Deposition: 2004/1/19
Taxonomy: Halobacterium salinarum
MMDB: 26602 PDB: 1S52 Related Structures: VAST
View options (Click image to view 3D structure)
Download Cn3D!
Molecular components in the MMDB structure are listed below. The icons indicate macromolecular chains, 3D domains, protein classifications and ligands. Please hold the mouse over each icon for more information on the component. You may also click the thumbnails below to view corresponding chains and domains in Cn3D.
Protein
3d Domains
Domain Family
Sequence A
Bac_rhodopsin
Protein
Domain Family
Sequence B
Bac_rhodopsin


NCBI
Related Structures
VAST
PubMed BLAST Structure Taxonomy OMIM Help? Cn3D
VAST related structures for: MMDB 26602, 1S52 sequence A.
Overview: There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced related structure search controls. The second section is the VAST related structure list itself.
View 3D Alignment of All Atoms with Cn3D Display Download Cn3D!
View Sequence Alignment using Hypertext for Selected VAST related structures
List Medium redundancy subset, sorted by Aligned Length in Table
Graphics
Table
Download Asn1
Download Xml
Entrez
Advanced related structure search
Move the mouse over the red alignment footprints in the graphics below and click, you will obtain a structure-based sequence alignment.
Total related structures: 165; 26 representatives from the Medium redundancy subset displayed.
Click to: Check All Uncheck All
1S52 A
3d Dom.
Protein Family
Ali_len
Chain A
Bac_rhodopsin
1VGO B 227
1C3W A 222
1E12 A 220
1H2S A 216
1XIO A 208
1JFP A 170
1OW8 C 135
2P7V A 120



NCBI

Related Structures
VAST

PubMed | BLAST | Structure | Taxonomy | OMIM | Help? | Cn3D

VAST related structures for: **MMDB 26602, 1S52 sequence A**.

Overview: There are two main sections to this page. The first section consists of the alignment view controls, the list controls, and the advanced related structure search controls. The second section is the VAST related structure list itself.

View 3D Alignment of All Atoms with Cn3D Display *Download Cn3D!*

View Sequence Alignment using Hypertext for Selected VAST related structures

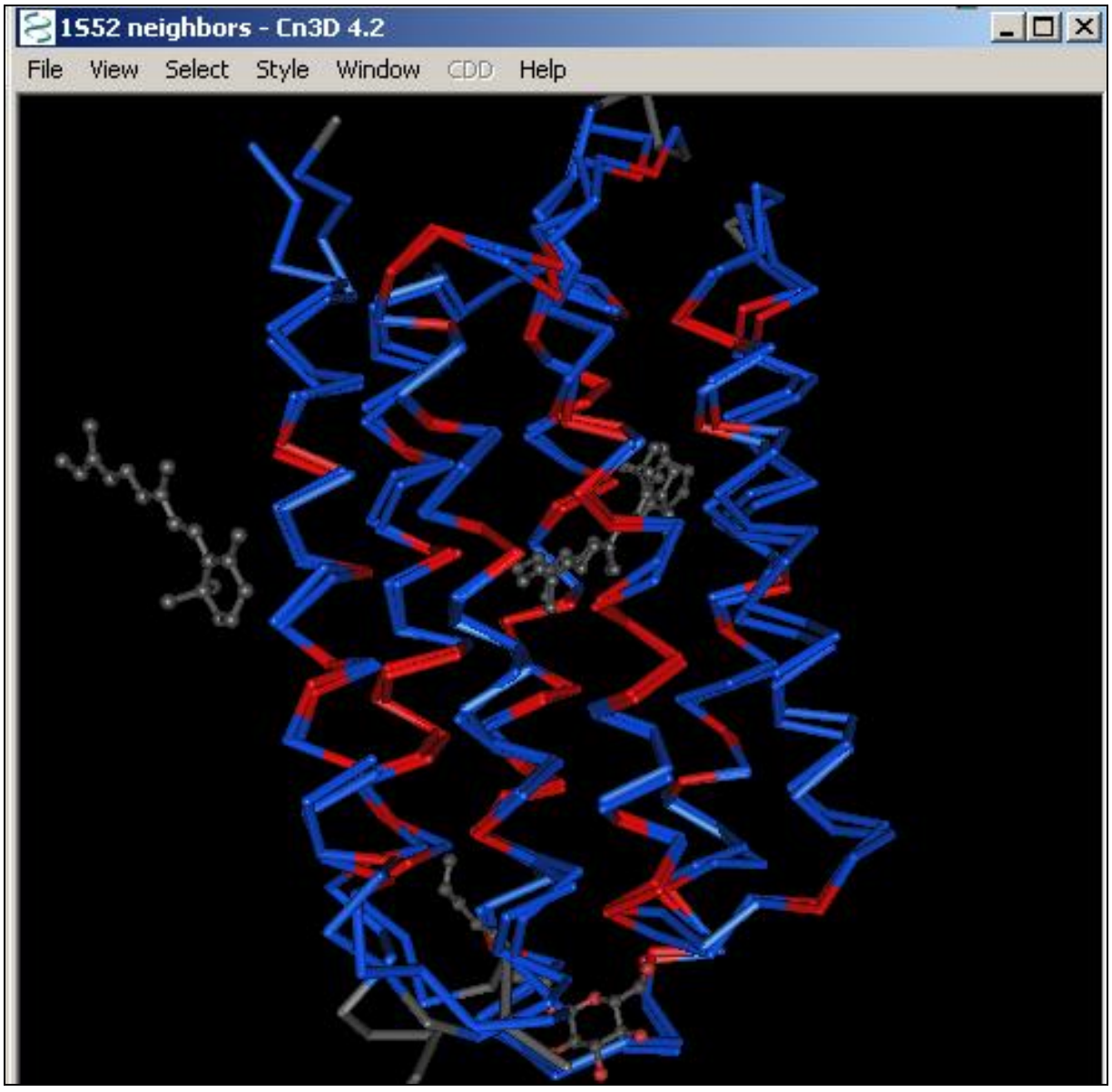
List Medium redundancy subset, sorted by Aligned Length in Table

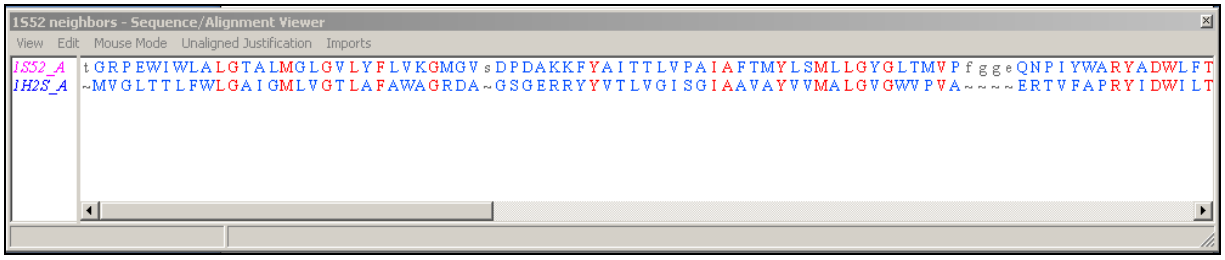
Advanced related structure search

Total related structures: 165; 26 representatives from the Medium redundancy subset displayed.

Click to: Check All Uncheck All

	PDB C D	Ali. Len	Score	E_Val	Rmsd	%Id	MMDB Date	LHM	GSP	Description
☐	1VGO B	227	15.7	10e-16.3	1.1	56.4	10/2005	0.3	0.5	Crystal Structure Of Archaerhodopsin-2ÿ
☐	1C3W A	222	16.0	10e-17.3	0.8	99.5	03/2001	0.8	0.4	BacteriorhodopsinLIPID COMPLEX AT 1.55 A RESOLUTION
☐	1E12 A	220	15.2	10e-15.1	1.5	34.1	03/2001	2.2	0.7	Halorhodopsin, A Light-Driven Chloride Pump
☑	1H2S A	216	15.3	10e-15.2	1.1	28.7	11/2002	1.5	0.5	Molecular Basis Of Transmenbrane Signalling By Sensory Rhodopsin Ii-Transducer Complex
☐	1XIO A	208	11.4	10e-9.8	1.6	29.3	11/2004	2.5	0.8	Anabaena Sensory Rhodopsin
☐	1JFP A	170	6.7	0.0351	3.9	8.8	11/2001	10.8	2.4	Structure Of Bovine Rhodopsin (Dark Adapted)
☐	1OW8 C	135	4.8	0.0353	4.4	7.4	11/2003	21.0	3.3	Paxillin Ld2 Motif Bound To The Focal Adhesion Targeting (Fat) Domain Of The Focal Adhesion Kinase
☐	2P7V A	120	5.7	0.0372	2.7	8.3	11/2007	NA	2.3	Crystal Structure Of The Escherichia Coli Regulator Of Sigma 70, Rsd, In Complex With


1S52 neighbors - Cn3D 4.2
File View Select Style Window CDD Help


1S52 neighbors - Sequence/Alignment Viewer
View Edit Mouse Mode Unaligned Justification Imports
1S52_A tGRPEWIWLALGTALMGLGVLYFLVKGMGVsDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVPfggeQNPIYWARYADWLFT
1H2S_A ~MVGLTTLFWLGAIGMLVGTLAFAWAGRDA~GSGERRYYVTLVGISGIAAVAYVVMALGVGWVPVA~~~~ERTVFAPRYIDWILT

Problem 2

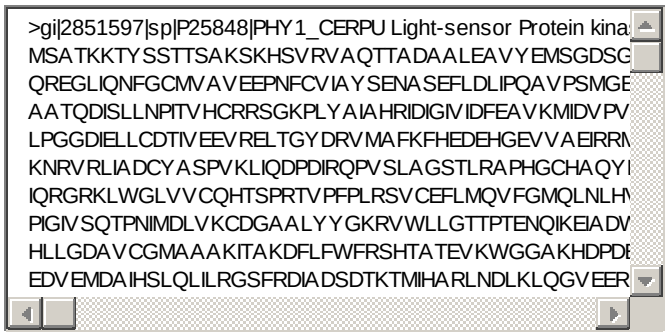
In this problem, we will follow these steps:

A. Identify conserved domain(s) present in a protein.
B. Search for other proteins containing similar domain(s).
C. Explore a 3D modeling template for the query sequence.
D. Find distant sequence homologs that may not be identified by BLAST.

NCBI's Conserved Domain Search allows you to match your protein sequence to a library of conserved protein domains, generate a multiple sequence alignment based on this match, and explore 3D modeling templates for your sequence. Click on the CDD link provided below,

CDD

paste the following protein sequence in the CD-Search query box and run the search.



A. What are the domains present in this protein?
(Select the “Full Result” radio button to display all of the domains.)

-Suppose, we are interested in the serine/threonine protein kinase domain. Obtain more information about it by searching in NCBI's Bookshelf

B. Go back to the CD-Search results page. Obtain a list of proteins with similar domain architecture by clicking on the "Search for similar domains architectures" button. To display the records, click on the links to the subsets of sequences and from there on the “Look up Sequences in Entrez”. Change the display from “Summary” to “FASTA”.

C. Go back to the CD-Search results page. Click on the “Full Report” radio button. Generate a multiple sequence alignment for the top 10 sequences representative of the conserved domain hit by clicking on the graphic representation of the serine/threonine kinase domain from CDD (CDD|00180). Use the "Aligned Rows" list box pull down menu to specify "up to 5" sequences

and invoke Cn3D with a display of a 3D modeling template and a multiple sequence alignment including your query sequence by pressing the "Show Structure" button.

To show only one top structure, click on the down arrow key. For better view of the backbone, remove the side chains globally (Style--Edit global style--Protein side chains). The query protein contains a serine/threonine protein kinase active-site signature (IIHRDLKSMNILV) where K is the ATP binding site. Identify these residues in the query protein and highlight the corresponding lysine residue in the first protein sequence.

Display the side chains of this residue (Use Style--Annotate--New--Edit Style. Change the protein backbone Rendering to Tubes, Color Scheme to User Selection and User Color to choose the color for the highlighted residue, for example yellow. Repeat these steps for the Protein Side chains row and click the Protein Side chains on. Click on the “Done” button. To zoom in, press z on the keyboard. Note the heterogen near the conserved lysine residue.

D. To obtain the structural neighbors for the serine/threonine protein kinase protein, first click on the structure entry link 1JNK of the similar protein from the CD-Browser page. Type 1JNK_A in the search box and click on the Go button. Then click on the structure link on the top right side, then on 1JNK, and finally on the chain graphic. Select one or more of the check boxes next to the structure neighbors and download the structures by clicking on the "View 3D Alignment" button.